

汉语口语成绩测试研究综述

梁 冬 梅

口语测试是语言测试中的一个重要组成部分,用来测定受试者运用语言进行口头交际的能力。汉语口语测试是考察母语非汉语的学习者汉语水平经常采用的重要形式之一。按照考试的目的分为:学能测试、分班考试、成绩考试、水平考试。汉语口试的整个环节包括:命题设计、测试实施、口试评分和成绩分析。近年来对汉语口语测试的研究主要集中在口试的题型及测试任务的难度分析;施测过程中,考官的提问技巧,及可理解性输入;口试信度研究、评分标准、评分方法、流利性研究等。对学能测试、分班考试的研究相对不足,对水平考试及成绩测试的研究较多。对汉语口语成绩测试研究重要的课题包括:测试的原则、怎么考(考试形式)、考什么(考试题型、考试任务及难度影响因素研究)、怎么评分(评分标准)等。

一. 口语成绩测试的原则

在汉语作为第二语言的教学,口语教学对培养学生的口语交际能力具有重要作用,是主干课程之一。如何准确评价学生的口语水平、评估口语课的教学效果,离不开口语成绩测试,因此成绩测试与日常口语教学的关系最为密切。

成绩测试包括在不同教学阶段进行的各种测试,如单元测验、月考、期中考试、期末考试等,它直接反映学生在一个阶段内的学习成绩和教师的教学效果。具有强制性、经常性和反馈性。其具体功能包括:1.诊断教学,具有评价功能;2.规范教学,具有导向功能。3.催化教学,具有激励功能;

4. 促进教学，具有反拨效应；5. 搜集信息、研究教学。¹

关于成绩测试的原则，彭国庆、贾磊提出了：1. 一致性原则。即测试的目标内容和教学的目标内容是一致的。2. 可靠性原则。成绩测试能稳定可靠地反应学习者的实际水平，即测试要具有可信度。3. 有效性原则。测试的内容和方法能有效地反应学习者的实际水平，即测试要具有效度。4. 区分性原则，指测试应具有区分学习者知识和能力差别的性能，即试卷的难易度要适中，试题间的难易程度要有跨度和一定比例。5. 操作性原则，是指测试要经济，便于施测，操作性强。特别是口试的形式决定了它有别于笔试，在实施测试的过程中，更要求经济，易于操作，特别是评分标准要具体，操作容易²。前四项都是一般成绩测试的基本原则，第五项原则结合了口语测试的特殊性。

许希阳针对汉语口语测试提出了“以学生为中心”的测试理念，认为口语测试的目的是复现学生真实的口语水平，因此应该“以学生为中心”。提出了真实性和交互性原则。1. 真实性原则：指受试者在测试中使用目标语完成测试任务与其在现实生活中使用目标语进行交际活动的相似程度。包括：输入的真实性（测试文本、测试任务），输出的真实性和情景的真实性。2. 互动性原则：体现在考生与考试任务之间的互动性以及考生与主考教师之间的互动性。“以考生为中心”，就是让考生有话可说而且愿意表达。如考生熟悉的、适合口语水平的话题、轻松自然的交际场景等。教师在场景中充当交际的另一方，与学生展开平等的对话，通过提示、引导、启发等方式让学生充分展示语言水平，在这种情况下才实现了考生与教师之间的互动性。³

许提出的原则针对一般的口语测试，没有考虑口语成绩测试的特殊性。但这些原则可以作为设计口语成绩试题及考试形式时的重要参考。

结合已有的研究及口语成绩测试的特殊性，我们认为汉语口语成绩测试的基本原则应包括：一致性、有效性、区分性、操作性和交际性。

二. 怎么考——汉语口语成绩测试的形式

柴省三、杨翼、王晓君等对汉语口语测试的形式进行了研究。按照考试

的形式，口试分为：间接口试、直接口试和半直接口试。在汉语口语教学班级中，学生一般不超过20人，因此汉语口语成绩测试大部分采用的是直接口试的方法，即考官与考生面对面进行交流，随时观察、记录，考完就给出成绩。测试人员往往是口语教师本人。直接口试交互性、真实性强。考完就给出成绩，所以效率比较高。不过，在直接式口语测试中，因为考官要控制考试程序、扮演交际角色、记录学生表现，又及时给出学生成绩，所以对考官的现场反应能力及速度要求比较高⁴。如何在较短的测试时间内对测试对象的口语水平进行准确评判，需要测试人员有极为灵敏的反应能力，稍有不慎，即容易造成误判⁵。考官即任课教师对考生的口语水平比较了解，评分时不一定完全根据考生现场表现，会受先入为主的印象影响而调整成绩，主观性较强，这也会影响测试的信度。

也有采用考官与考生面对面进行交流、录音，考试后听录音评分，考试者和评分者可能由不同的人担任，比如由平行班的不同任课老师打分。这种形式提高了评分的准确性，但比较花费时间和精力。

也有人提出在口语成绩测试中采用标准化试题，让参加测试的留学生进行口语录音，考官只负责控制录音设备。录音结束后，教师听录音，给出成绩。认为这种测试方法具有较高的可靠性、稳定性和统一性，避免了过去那种单凭印象打分产生的偏差，有助于教师对考生成绩的定位分析，也有助于教师对学生作测试后的指导，可充分地发挥测试后的反馈效应⁶。但这种测试形式失去了口语考试的真实性和互动性，似乎更适合大型的标准化的口语测试，如HSK口试，是否能作为理想的口语成绩测试形式有待商榷。

随着口语测试领域研究和科技的进步，更多的口语测试形式被开发出来，如电话口试和人机互动的口试。但是如何把这些新形式更好地应用到口语成绩测试中将是新的研究课题。

三、考什么——汉语口语成绩测试的任务

1. 口语成绩测试的题型

考试的题型取决于考试的目的及可行性。学者们根据考试目的不同，研

究了考试的题型。李海燕等(2003)等研究了口语分班测试题型。杨翼(2014)、彭国庆、贾磊(2012)、许希阳(2005)、陈昭玲(1988)、翟艳(2012)等关注了口语的成绩测试的考试题型。

汉语成绩口试的内容范围就是教学大纲规定的教学项目和教材的内容。口试的目的是检测学习者用汉语进行口头表达的能力。即能否用正确的语音、语调得体、流利地交流和表达,在表达过程中,词语使用是否丰富、恰当,语气是否自然;语法结构、口语句式是否正确;能否根据交际的需要灵活选择和变换不同的说法,表达是否流畅,是否能达到交际的目的。鉴于此,汉语口试的内容包语音能力(声、韵、调、停顿、重音、语调、语气等)、用词造句能力、成段表达能力、语用能力、跨文化的口语交际能力等。⁷

口语成绩测试,具有检验学习内容和诊断口语水平的双重职责。因此既需要考察交际性,又需要考察学生对学习内容和语言形式的掌握情况。成绩测试也带有明显的阶段性。考试只是对目前这个学习阶段学生口语水平的认定,而且对口语能力诊断的意义大于对口语能力判断的意义。⁸

鉴于汉语口语成绩考试的目的和特点,试题应当包括主观性试题和客观性试题,分立式试题和综合性试题。

杨翼(2014)概括了几种常用的汉语口语成绩测试的题型:⁹

- (1) 朗读:包括拼音、汉字、单词词组、句子语篇等,考察语音、语调、重音、连读、节奏、停顿、语流音变等。
- (2) 复述:考生先阅读材料,然后复述。许希阳提出:教师应参与其中,让学生看材料,然后根据所看内容和教师一起对话,对话既要呈现材料中一些基本、重要的信息,而且随着对话的展开学生可以即席加入内容。
- (3) 问答:包括被试者回答考官提出的问题和被试者朗读一段语料后回答问题。许希阳建议教师与学生进行问答时,应把大的话题切分成几个子话题,按照逻辑顺序层层展开,通过学生的回答获得信息,进行真实、互动的信息交流。避免让学生回答没有逻辑关系的几个问题。¹⁰
- (4) 角色扮演:规定大致情节及交际场景,为考生设定角色,进行师生之间、生生之间的对话。交际性口试特别推崇的题型。

(5) 看图说话：提供图画或者照片，要求学生回答问题（或描述景物、人，或叙述事件）；或者用图片引出话题。能引发学习者的口语表现，又能较好的避免阅读或听力技能对口语表现的影响。

(6) 口头演讲：用汉语发表意见、论证观点、解释原因、讲解办法。先提供题目或一段材料，准备后再口试。应该是学习者感兴趣的事件或者问题。

许希阳提出成段表达事先让学生准备无可厚非，但学生几乎原封不动地背诵就不符合口语行为的即席性特征。应该要求学生在新的情境下对原先的东西进行重新编码和加工。¹¹

(7) 小组讨论：3、4个学生根据给定的话题进行口头交流，以观察学生的汉语口语常规技能（声韵调、流利度和正确度）和汉语应变能力。话题应新颖有趣，应考察学生如何引入话题、阐述思想、如何结束、表达的清晰度、语言组织的逻辑性、说话时长等。

(8) 口译：汉语翻译成外语，外语翻译成汉语。

翟艳（2012）通过调查整理出了4个单位或文献中的口语成绩测试的题型，如下图。¹²

表1 题型统计

题型	条数	来源 [®]	题型	条数	来源	题型	条数	来源
造句	14	4	看图说话	4	3	选择判断	2	1
朗读	14	4	解释词语	4	2	演讲	2	1
回答问题	12	4	角色扮演	3	2	说词	1	1
话题表述	9	4	辩论讨论	3	3	听写	1	1
复述	8	4	讲故事	2	2			
对话	7	3	小组调查	2	1			

许希阳认为在设计口语试卷时应该摆脱练习中存在的机械、人为的痕迹，创造典型、真实的情境让学生能够较为自然地运用所学的内容。认为完成句子、造句、改说句子、解释词语等不太合适。

翟艳认为朗读、造句、复述、解释词语、选择判断、说词与听写还停留在结构主义的语言形式考察上。多是单向、静态的表达,学生在给定的内容框架下作答,内容受限制,机械性的表达居多,完全不具有任务型测试特征。她结合任务型汉语教学提出以被试完成任务的表现来评估学生的语言能力。测试过程采用动态的一对一或小组互动的方式,通过信息差等任务活动,激发学习者的主动参与意识和表达。提出任务型汉语口语成绩测试题型:1. 看图说话;2. 听指令,做动作,同时做出语言回应;3. 角色扮演——考生根据情境和提示,轮流向对方提问,进行角色扮演。或是小组调查——一是提前布置任务,考试时,学生以ppt形式展现报告的结果。一种是现场调查,考生随机分组,根据主题进行组内调查,记录要点,现场公布调查结果。¹³

我们认为,结合成绩测试的原则与功能,机械的、侧重语言形式的测试题型是必要的,但是所占比例不应太大,如不超过1/3,大部分题型应该是接近真实交际情境的试题形式。

2. 口语测试任务难度的影响因素

不同的口语测试题型难度是否相同,影响难度的因素有哪些?(聂丹, 2011, 2012)对测试任务难度的影响因素做了系列研究。

口语测试的任务,就是在测试环境下要求考生使用目标语口头完成测试作业或试题,其功能是引导考生说出可供评分的话语,以此推论考生在目标语交际环境中的口语能力。作为口语测试的重要基础,任务及任务难度关系到口语测试的效度和信度,其重要性不言而喻。

通过研究,她发现:对被试口语表现影响程度由小到大的因素依次是:时间 < 环境 < 考官 < 交际任务 < 话题内容 < 考生心理状态 < 考生口语能力。可见,考生个人的能力和状态是影响测试表现的主要因素。叙述、描述和议论说理这三种任务类型由易到难依次为:叙述任务(介绍一个人或讲述一件事)< 描述任务(看图说话)< 议论说理任务(讨论一个问题或评说一个道理)。对任务难度的影响程度从大到小,从大类来看,依次是:话题因素 > 输入因素 > 组织与环境因素 > 语言因素 > 体裁因素 > 被试因素。

从小类来看,依次是:话题熟悉度 > 词汇难度 > 话题类型 > 输入语言的难度 > 输入问题理解度 > 话题认知度/输入方式接受度 > 活动方式和类型 > 测试环境 > 体裁类型 > 话题抽象度 > 输入情境或提示信息的激发性 > 话题相关度 > 语法句法难度 > 体裁组织思维复杂度。口语考试的三种组织形式:考生独白、考生与考生对话、考生与考官交谈,从考生不喜欢到喜欢的顺序是:独白 < 考生一考生 < 考生一考官。从考生不紧张到紧张顺序依次是:考生一考生 < 独白 < 考生一考官。按照考试有效程度从低到高依次是:考生一考生 < 独白 < 考生一考官。考官与考生面谈这种形式让考生感到紧张,但却是最受考生欢迎的,同时被认为是最有效的。¹⁴

通过对体裁难度的感知调查和实测调查,她发现体裁认知类型由易到难的序列:叙述、描述 < 说明——类举、说明——比较 < 说明——阐释 < 议论——评测 < 议论——论说、说明——决策。¹⁵

以上的研究对结合考试任务难易程度设计合理的测试任务、测试题型具有重要意义。考试任务对不同背景、不同特点的考生是否具有不同的难易程度,目前这方面的研究还不足。

四. 口试的评分标准

口试的成绩评定,涉及到评分标准和考官或者是评分员两种因素。鉴于汉语口语成绩测试基本是由任课老师一人担任评分员,我们暂且只关注了对口语评分标准的研究,暂不考虑对评分员的研究。

从哪些方面判断口语水平,即口试的评分标准是什么?一直是口试研究的重点课题之一。近些年来的测试研究,特别是大型的口语测试研究,都较关注评分标准的讨论。也都有一套与测试概念相适应的评分标准。英国的IELTS:流利性和连贯性、词汇、语法及语音四个维度。美国ETS的TSE口试:语法、语音、流利性和可理解性四个方面。FSI口试(Foreign Service Institute):语音、语法、流利性、词汇和可理解性五个方面。而中国大学英语口试SET(Spoken English Test):准确性、语言范围、话语长短、连贯性、灵活性和适切性六个方面。汉语口语考试(HSKK),陈田顺从内容、语

音语调、流利程度、词汇、语法等方面分解了标准。刘镰力认为考官应首先重视总体印象,在此基础上注意对内容、语音语调、语法、得体性这几方面的把握。美国的OPI(Oral Proficiency Interview)评分时考虑了功能 / 情景(即交际场合)、话题(谈话的广度、准确性)、篇章类型(输出的是段落、句子还是单词、短语)四个方面。¹⁶显然,对于评价口语水平的基本维度究竟有多少、不同维度之间的关系如何以及不同维度在口试分数中所占的权重分别有多大等问题,在语言学界还没有一个公认的标准。

规范化、标准化程度都不高的口语成绩测试其主观性更强,影响评估信度的关键项目——评分标准是一个从研究到操作都不成熟的薄弱方面。评分员一般都由任课教师担任,考试的随意性、标准的不确定性和对标准认识的模糊性现象更为严重。目前各校口语成绩测试多为教师一人打分,评分方法主要为按照不同等级的相关标准、利用打分表进行分项打分。实际上,在具体操作中,据教师反映,由于比较熟悉学生,老师们在心里多有“先入之见”,如果学生的表现失常,在分项评分后都会有一次“位置调整”,这时起作用的就是综合评分。因此翟艳建议任务型口语成绩测试的评分方法采用分项评分与总体评分相结合的方式,以分项评分为主,总体评分为辅,总体评分的分值设为总分的5%-10%。¹⁷

许希阳(2005):期中、期末口语测试时,分数评定基本上有两类:一种是教师按学生每个题目的回答情况打分,最后计算总分;另一种是教师根据学生的表现给一个总的印象分。前者看似公平,但只有评定学生对语言知识的掌握才能保证公平;后者是一种综合的印象式评分,没有明确的标准,主观性太强。口语测试更强调交际的原则,因此不应只注重学生话语的准确性,即是否有语音、词汇、语法错误,而且应该注重实际表达的内容所体现的意思,以及话语的流利性、得体性、连贯性、逻辑性、完整性、思辨性等各种要素。此外,口语成绩用百分制评定很难准确地说明学生口语水平的高低。因此她建议列出具体评估项目,并给出成绩级别。

表2

	语音	词汇	语法	内容	流利性	连贯性	得体性	完整性	整体交 际效果	其他
优										
良										
一般										
及格										
不及格										
总评										

初、中、高不同的等级考察学生口语能力时侧重点也各不相同，因此评定学生最后的口语成绩应该着眼那些最关键的指标，以此凸显不同等级的特殊要求。学生拿到成绩单时不仅看到一个总的评价，也看到了这些具体项目的评定情况，这对他们进一步提高口语水平有着借鉴的意义。¹⁸

张军（2014）调查了对外汉语教师心中的口语评价标准，得出的结论认为：在“语音（语音语调的准确性、可理解性）、词汇（词汇的广度、选词的恰切性）、语法（语法的准确性、复杂性）、流利度（语速、节奏和韵律、非正常停顿——如长时间搜索词汇、思考、自我修正、重复）和内容（完整性、连贯性、逻辑性、关联词的使用）”五项标准不是同等重要的，在各个教学阶段中其所占的比重有较大不同。留学生口语产出的语音、词汇和语法三项是教师更为关心的，特别是零起点的汉语教学，语音项权重占到三分之二。随着教学活动的开展和学生汉语水平的提高，各项标准有均衡发展的趋势，但始终是语音、词汇和语法三项为主，流利度和内容为辅。¹⁹

杨桦（2013）认为：高级口语课的成绩测试，应以下面十个方面作为标准进行评判：充分的准备；发音和语法比较规范；时间长度合适；观点清晰；内容充实；结构完整；表达流畅；逻辑合理；语句通顺连贯；使用适当的表达用法和词语。这样既可以充分体现高级口语课的训练目的，又可以切实提高学生的高级口语交际水平，同时在相当程度上保证在评判标准方面做到细致统一。²⁰

翟艳（2012）认为考试标准与一定的任务形式相联系，口语评估标准要做到简洁、易行、区分清晰，她探讨了针对成段表达任务设立的测试标准。确立了五大项 11 子项的口语成绩测试标准，如下图²¹

表3 五项标准的子项

项目 数量	听感			语言形式					意义表达		
				词汇		语音	句法				
	停顿少	语流自然	重复少	使用正确	词汇丰富	语音语调正确	使用正确	句式复杂丰富	表达正确完整	衔接连贯	逻辑清晰

口试评分标准因其重要性和复杂性，引起了广泛的关注，形成了一定的共识：如语言形式（语音、词汇、语法—正确性，丰富性）、流利度（或听感）、意义表达（正确、连贯、逻辑清晰）等，也有学者进一步关注了流利度的研究。但是每一项标准的等级如何制定，每一项标准在总成绩中应占多少比例，在不同教学阶段不同标准具体如何侧重等，还没有看到更多的研究成果。汉语口语成绩测试亟需确立更科学的、可操作的统一的评分标准。

¹ 彭庆国、贾磊，汉语成绩测试与命题设计——以汉语口试为例，和田师范专科学校学报，2012年11月，20-21

² 同上，21-22

³ 许希阳，汉语口语测试研究，云南师范大学学报，（对外汉语教学与研究版），2005年9月，第3卷第5期，40

⁴ 柴省三，汉语水平口试信度的理论与实证研究，语言教学与研究 2003年第4 期，70

⁵ 王晓君，对外汉语口语成绩测试形式试探，首都师范大学学报-社会科学版，2013年增刊，108

⁶ 张慧君，对外汉语口语课的成绩测试创新，教育探索，2006年第7期，9

⁷ 杨翼，对外汉语教学的成绩测试，北京大学出版社，2014

⁸ 翟 艳，汉语口语成绩测试评估标准，华文教学与研究，2012年第1期，总第45

期, 43

⁹ 同7

¹⁰ 许希阳, 对外汉语口语成绩测试新模式之探索, 语言教学与研究, 2008年第4期, 78

¹¹ 同10

¹² 翟艳, 任务型汉语口语成绩测试研究, 语言文字应用, 2012年11月第4期, 119

¹³ 同上

¹⁴ 聂丹, 对口语测试任务难度影响因素的调查研究, 中国考试, 2011年第12期, 27-29

¹⁵ 聂丹, 汉语口语测试任务的体裁难度评估, 语言教学与研究, 2012年第3期, 49

¹⁶ 翟艳, 汉语口语成绩测试评估标准, 华文教学与研究, 2012年第1期, 总第45期, 43

¹⁷ 同上

¹⁸ 许希阳, 汉语口语测试研究, 云南师范大学学报-(对外汉语教学与研究版), 2005年9月, 第3卷第5期, 45

¹⁹ 张军, 高校初级汉语口语评价标准的调查研究, 考试研究, 2014年第4期, 总第45期, 48

²⁰ 杨桦, 高级口语成绩测试评判标准研究, 首都师范大学学报-社会科学版, 2013年增刊, 140

²¹ 同16, 48

其他参考文献:

1. 陈昭玲, 大型对外汉语口语成绩测试的探索, 汉语学习, 1988年第4期, 总106.
2. 邓秀均、丁安琪, 论对外汉语口语测试中的提问技巧, 对外汉语研究, 云南师范大学学报, 2003年5月第1卷第2期, 1-5
3. 李梦莉, 口语测试交互性研究, 长春大学学报, 19卷第9期, 2009年9月, 101-107
4. 梁冬梅, 留学生入学分班汉语口试监考教师语言分析, 人文丛刊, 2014年6月, 第八辑, 157-167
5. 聂丹, 面试型汉语口语测试的可理解性输入探析, 汉语学习, 2011年04月, 第2期, 75-82
6. 郭修敏, 汉语作为第二语言的口语流利性的量化测评, 北京语言大学硕士研究

生论文，指导教师郭树军，2005年6月

7. 李海燕蔡云凌刘颂浩，口语分班测试题型研究，世界汉语教学，2003年第四期，79-89
8. 刘晓雨，对外汉语口语教学研究综述，语言教学与研究2001年第2期，27-32
9. 赵琪凤，汉语口语测试的信度研究与教学启示，汉语学习，2012年10月，第6期，89-94
10. 张斐斐，汉语口语测试评分信度探究，中山大学学报论丛，2007年第27卷第11期
11. 张和生OPI与汉语口语水平的测试，北京师范大学学报（社会科学版）1996年第6期（总第138期），96-100